DOCUMENT RESUME

ED 204 395 TM 810 390

AUTHOR Koch, William R.
TITLE Dimensions Underlying Student Ratings of Instruction.
PUB DATE Apr 81
NOTE 18p.: Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Course Evaluation; *Factor Analysis; *Graduate Students; Higher Education; *Multidimensional Scaling; Rating Scales; *Student Evaluation of Teacher Performance; Undergraduate Students

ABSTRACT

A study was conducted to determine the dimensions underlying the items on a course-instructor survey instrument currently in use to measure graduate student ratings of instruction (attached). Both factor analysis and multidimensional scaling methods were used to analyze the response data, so that the results of each could be evluated, and the two techniques could be compared. The results indicated that the dimensions salient to graduate students in ratings of instruction were similar to those commonly reported for undergraduates. However, the multidimensional scaling procedure was found to have several advantages over factor analysis, especially in ease of interpretability. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Koch, W.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED204395

DIMENSIONS UNDERLYING STUDENT RATINGS OF INSTRUCTION

William R. Koch
University of Texas at Austin

Reviews of the literature in the general area of course and instructor evaluation (e.g. Kulik & McKeachie, 1975) have shown that certain factors consistently emerge from analyses of student ratings of instruction. For example, it has been frequently observed that a distinction can be drawn between those items on student-completed rating scales that address the course and those that address the instructor. On a more specific level, factor analytic studies have identified a common core of factors that recur, although the verbal labels assigned may change. Such factors as the skill of the instructor, the rapport between the instructor and the students, the organizational structure of the course, the amount of work demanded by the course (difficulty), the level of interaction between the instructor and students, and the amount of feedback given to students have often been reported.

Most of the studies in the literature have been restricted to the ratings of instruction given by undergraduate students, rather than graduate or professional area students. Also, the findings summarized above have typically resulted from data analysis techniques such as traditional item analysis, factor analysis, or discriminant function analysis.

The present study was motivated by two main questions: a) In analyzing graduate student ratings of instruction, would the same general dimensions emerge as have been found for undergraduate students?, and b) Might there be some advantages and/or differences in the dimensions that result when the data are analyzed using multidimensional scaling instead of factor analysis?

TM 810390

Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, 1981.

Starting with the first question, it seems that on the surface one might expect that graduate students would differ from undergraduates in the importance assigned to certain aspects of the course or the instructor. Because the class size is usually smaller in graduate courses, the students might reasonably expect to develop close rapport with the instructor. Also, graduate students might place more importance on the scholarly reputation of the instructor or on the relevance of the course content to career objectives. Furthermore, aspects of graduate courses such as workload, outside reading assignments, research paper requirements, and exams often differ considerably from their counterparts in undergraduate courses. Thus, the dimensions underlying student ratings of graduate instruction could be quite different from those found for undergraduate instruction.

Taking the question of multidimensional scaling versus factor analysis to analyze the rating data, several differences in the assumptions and the methodology of these two techniques could lead to variations in their results. For example, factor analysis assumes that the item scores are measured on an equal interval scale, while multidimensional scaling only assumes ordinal measurement scales. Also, factor analysis is a vector model based upon assumed linear relationships among the variables, while multidimensional scaling is a spatial model based upon the weaker assumption of monotonic relations among the variables. Another common difference is that multidimensional scaling (MDS) usually results in a lower dimensional solution (fewer dimensions) than does factor analysis of a given set of data. It is frequently an advantage to be able to graphically present the two- or three-dimensional configuration resulting from MDS, while the higher dimensional solution from factor analysis cannot be visualized directly. (See Kruskal & Wish, 1978, and Shepard, Romney, & Nerlove, 1972, for discussions

of multidimensional scaling analysis.) Because of the simple structure of many MDS solutions, it is often relatively easy to interpret the results. Another advantage of MDS is that interpretations may be made on the basis of items that cluster together in the configuration, rather than strictly on the basis of dimensional axes in the space.

METHOD

Instrumentation

The rating scale analyzed in the present study was the Educational Psychology form used in the Course-Instructor Survey program at the University of Texas at Austin. Like the other 21 separate questionnaires currently in use for the Course-Instructor Survey (CIS) program, the form had been constructed by a teaching effectiveness committee within the Educational Psychology Department. The questionnaire was designed to measure six specific areas of instruction, including: a) knowledge of subject matter, b) organization and presentation of materials, c) instructor-student interaction, d) instructor enthusiasm and confidence, e) contribution to motivation and knowledge, and e) handling of instructional tasks. Also, there was a series of items to collect general demographic information, as well as a set of four global items about the course and the instructor. A copy of the questionnaire is included in Appendix A of this paper. Because the great majority of courses taught in the Educational Psychology Department at UT-Austin are graduate level courses, the items on the questionnaire were oriented toward graduate instruction. The total form had 40 items on it, but the demographic items were dropped, leaving 34 items to be subjected to the analyses of the present study. All 34 of the items included were scored on a five-point scale in typical Likert fashion.

Sample

The response data were collected as part of the routine administration of the CIS program near the end of the semester in Spring of 1980. A total of 663 students participated in the survey (some rating more than one course or instructor), rating 36 different instructors for 51 educational psychology graduate courses taught at the University of Texas. The data were collapsed for analysis. It should be noted that participation in the CIS program is voluntary for both faculty and students, with the instructor maintaining control over the release of the results obtained from each particular course. Due to selection effects from voluntary participation, the representativeness of the data is open to question.

Analyses

The basis for both the MDS analysis and the factor analysis of the response data collected on the questionnaire was the inter-item correlation matrix, since that matrix served as the input data for both procedures. The unit of analysis for the correlations was the mean response score to each item across the 51 classes. Several different factor analyses of the data were run using SPSS routines, including principal components analysis and iterative principal axis factor analysis, followed by both varimax and oblique rotations.

The MDS analyses were run using the ALSCAL program (Young, Takane, & Lewyckyj, 1978) within the Statistical Analysis System (SAS). Although the inter-item correlation matrix was input to ALSCAL, the MDS procedure did not use the actual coefficient values. Instead, only the rank orders of the magnitudes of the coefficients were used -- that is, ordinal scale data rather than interval scale data. MDS solutions were attempted in two, three, and four dimensions. Also, to aid in the interpretation of

the results, hierarchical cluster analyses were run using the SAS package. The coordinate values that located the items in multidimensional space (output from the MDS runs) served as input to the clustering procedure.

RESULTS

The preliminary results of the principal components analysis and the principal axis factor analysis indicated that only a single dominant factor was being measured by the instrument, with this factor having an eigenvalue equal to 20.89, accounting for about 61% of the total variance. Such a result is typical of scales that have been developed through item analysis to obtain high coefficient-alpha reliabilities reflecting internal consistency. However, in light of the purposes of the present study, the arbitrary rule of rotating all factors having eigenvalues greater than 1.0 was used with the data. The result was that five rotated factors were interpreted. It did not make any difference in terms of factor labels or items loading on factors whether principal components or principal axis was used, nor whether varimax or oblique rotations were performed, so only the results of the principal components-varimax combination are presented here.

Table 1 shows the loadings of each of the 34 items on each of the five components, or factors, which accounted for about 82% of the total variance. A subjective examination of the items loading on each factor resulted in the following factor labels:

factor I -- course aspects and value
factor II -- instructor attributes and lecture style
factor III -- rapport/interaction between instructor and students
factor IV -- relevance of course to the current field
factor V -- residual factor (no primary loadings)

The interested reader will want to refer to the actual items from the questionnaire listed in Appendix A to assess the reasonableness of the

factor interpretations given above. However, some examples are presented below to demonstrate typical items tha led to the factor labels.

Insert Table 1 about here

Some of the items loading on the course factor (factor I) included statements that the course increased the ability to evaluate work in the field, increased skills relevant to future work, stimulated outside reading, increased understanding, got the student interested, was valuable to the student, and was a good course overall. Factor II items, about the instructor and lecture style, consisted of statements that the instructor had thorough knowledge of the subject, the lectures were organized and clear, the instructor was self-confident, used examples and illustrations in lectures, was intellectually stimulating, and was a good instructor overall. The rapport/interaction items (factor III) stated that the instructor was aware of the class understanding, commented on written work, was accessible to students, respected the students, was aware of their needs and feelings, and related the class topics to their lives. The two items that loaded on factor IV were the statements that the course was related to current developments in the field and that the instructor could suggest outside reading sources.

The MDS procedures are a relatively new development in measurement compared to factor analysis, and they are only recently being routinely applied by practitioners as computer programs like ALSCAL become available. The basic goal of the MDS procedure is to achieve a one-to-one monotonic relationship between a matrix of stimulus similarities and the corresponding distances among the stimuli in a spatial configuration. This task is accomplished through an iterative process that attempts to minimize the disparities between the similarities and the distances, i.e. to minimize stress (Kruskal, 1964). If two objects are quite similar, they should also

be located quite close to each other in the spatial configuration, meaning that their coordinates should be similar in value.

For the present study, the inter-item correlations were arranged in rank order to serve as indicators of the similarity of one item to another. The three-dimensional solution was judged to be optimal because the stress had a reasonably low value of 0.13, while the MDS model accounted for 89% of the variance of the disparities. The two-dimensional solution had an unacceptably high stress value, but the four-dimensional solution did not improve much on the stress or variance accounted for by the three-dimensional configuration.

The subjective interpretation of the dimensions was accomplished by finding items at the extremes of each dimension. Table 2 presents the results of the three-dimensional MDS analysis, where the "endpoint" items are underlined. It appeared that all three of the dimensions defined bi-polar axes with item anchor points at either end, which represented opposite ends of the continuum for each dimension. Note that the scale of the coordinate values resembles that of standard scores rather than the zero to $\pm$1.0 scale used to indicate the loadings or correlations of the items with factors in factor analysis.

Insert Table 2 about here.

The first two dimensions measured aspects of the course and attributes of the instructor, respectively, just as did the first two factors discussed above. However, the MDS dimensions revealed a distinctive and interesting pattern, quite different from the factor analysis results. For the course dimension (I), one end of the continuum was defined by specific, concrete items about the exams, assignments, optional readings,etc., while the other extreme was defined by more subjective items such as depth of understanding,

freedom to ask questions or disagree, relation of subject matter to other areas, and relation of topics to students' lives. But note that the global item about the overall rating of the course was located closer to the concrete end of the dimension than to the subjective end, the implication being that more importance is placed on the practical aspects of the course than the enrichment aspects. Dimension II, dealing with the instructor, was anchored at one extreme by items about the general skill and lecturing style of the instructor (knowledge of the subject, organization of lectures, self-confidence, and pace), while the other extreme reflected the personal involvement of the instructor with the students (accessibility, feedback given, respect, etc.). The item asking for an overall rating of the instructor was located toward the skill end of the dimension rather than the personal involvement end. Dimension III appeared to be a type of course quality dimension, perhaps reflecting different needs of the students. Both ends of the dimension included items related to quality courses, but one end seemed to be more traditional (well-paced, good exams, relevancy,etc.) and the other seemed to be oriented more toward scholarly or research-type quality (currency, suggested readings for further study, and non-overlap of lectures and readings).

As was mentioned earlier, it is often desirable to interpret clusters of items located in the spatial configuration rather than having to "force" an interpretation based on axes (continua). The cluster analysis results for the present data are shown in Table 3. At each stage, the clustering procedure attempts to form groups of items such that the within group distances are minimized while the among groups distances are maximized.

The three cluster solution appeared to be optimal, the clusters consisting of items about the course characteristics (exams, readings,

assignments, relevance, overall course rating, etc.), items about the lecture style of the instructor (knowledge of the subject, organization and clear presentation, self-confidence, etc.), and items reflecting the attributes of the instructor and rapport with the students (enthusiasm, accessibility, respect for students, concern with teaching quality, intellectual stimulation, overall instructor rating, etc.). The four cluster solution also appeared to be satisfactory, the only difference being that four items (6,9,12,15) included in the third cluster now formed their own cluster. All four of the items dealt with interaction between the instructor and the students.

Insert Table 3 about here.

DISCUSSION AND CONCLUSIONS

In general the results demonstrated that approximately the same dimensions of instructional ratings were important to graduate students as have previously been found for undergraduate ratings of courses and their instructors. No comparisons were made in the present study of the relative importance placed on the dimensions underlying instructional ratings by graduate students versus undergraduates, however. The only dimension found in the present study that seemed somewhat unique was the third dimension in the MDS analysis. This was labeled a course quality axis, and it seemed to reflect different criteria of quality used by the students. At one end of the continuum were items reflecting highly organized courses (well-paced, good exams, relevancy), while the other end had items reflecting highly stimulating courses (currency, outside related readings, non-overlap of lectures and readings). Perhaps this dichotomy was related to the different needs of graduate students who were majoring in the course's field of study compared to those who were taking it as an elective (not directly related to their chosen field).

Other than this exception, all of the dimensions underlying ratings of instruction in the present study have been reported previously. It may be worth observing, however, that most of the questionnaires used for student evaluation of instruction are similarly constructed, often with very similar items. Thus, the consistency with which the same basic dimensions are "discovered" across studies may reflect the similarity of the questionnaires and not any "true" dimensions that underlie ratings. The counter argument is that the dimensions usually make sense intuitively.

In regard to comparisons of factor analytic techniques versus the MDS and cluster analysis approaches, no claim is made by the present study that one is in some sense better than the other. However, it is argued that the assumptions made by the factor analysis model are usually not met by typical rating scale data. In contrast, the weaker assumptions of the MDS model are invariably met. Moreover, the MDS results provided more insight to the resulting dimensions in terms of the ability to identify endpoints of bi-polar axes. To check on the possibility that the factor analysis might measure the same three dimensions that resulted from the MDS procedure, a separate analysis was run in which the three factor solution was specified (forced). However, there was not much agreement between the items with high loadings on the three factors and the items with extreme coordinates on the three dimensions.

In some ways, the cluster analysis results were the cleanest of all. Whether the three or four cluster solution was used, there was a nice correspondence between the mutually exclusive item groupings with what might be obtained through a subjective examination and sorting of items into groups based on their superficial similarities. Of course, the cluster analysis

procedure does not even assume that the items must have a dimensional or linear axis configuration, but only that groups may be formed based on similarities. Thus, the use of MDS, cluster analysis, or the combination of the two is recommended as a viable alternative to the traditional method of factor analysis to determine the dimensions or aspects of the items that underlie ratings of instruction.

TABLE 1

Principal Components Analysis
Varimax Rotated Factor Pattern

Item Number	Factors				
	I	II	III	IV	V
1	.27	.77	.18	.27	−.00
2	.34	.83	.13	.01	−.24
3	.29	.54	.71	.03	−.18
4	.51	.66	.43	.03	.09
5	.73	30	.44	.21	−.10
6	.10	.03	.78	.37	.19
7	.35	.48	.62	.18	.08
8	.49	.67	.42	−.01	−.02
9	.22	.34	.67	.22	−.22
10	.55	.53	.43	.15	.29
11	.81	.31	.32	.14	−.07
12	.28	.08	.75	.01	.29
13	.22	.59	.52	−.18	.12
14	.50	.61	.42	−.06	.05
15	.24	.39	.63	.38	−.09
16	.07	.85	.16	.21	.10
17	.85	.18	.26	.26	.13
18	.77	.26	.09	.24	.05
19	.30	.41	.32	.63	−.01
20	.57	.32	.44	−.05	−.26
21	.26	.37	.80	.16	−.08
22	.24	.68	.37	.03	.34
23	.71	.47	.42	.13	.05
24	.65	.16	.14	.09	.53
25	.49	.02	.09	.78	.13
26	.29	.59	.11	.27	.50
27	.15	.28	.65	−.23	.24
28	.41	.57	.56	.03	.07
29	.51	.64	.39	.22	.14
30	.52	.26	.57	.25	−.13
31	.52	.62	.46	.12	.07
32	.87	.28	.25	.15	.09
33	.50	.73	.34	.05	.12
34	.82	.35	.13	.08	.24

NOTE: The underlined values indicate the highest loading of an item on a factor. Broken underlines indicate other high loadings.

TABLE 2

Multidimensional Scaling Analysis
Item Coordinates for Three-Dimensional Configuration

Item Number	Dimensions I	II	III
1	.07	−1.42	1.13
2	.59	−1.98	.11
3	1.37	.20	− .01
4	.16	− .43	− .10
5	− .73	.63	− .20
6	.80	2.52	.85
7	.60	.46	.11
8	.36	− .58	− .46
9	1.45	1.07	.81
10	− .28	− .16	− .12
11	−1.16	.07	− .38
12	.68	2.16	− .99
13	1.69	− .69	− .90
14	.46	− .61	− .58
15	.67	.99	1.17
16	1.11	−1.82	1.12
17	−1.59	.56	− .07
18	−2.01	− .55	− .41
19	− .48	.34	1.84
20	.06	.18	−1.82
21	1.15	.94	.12
22	1.13	−1.07	− .06
23	− .35	− .04	− .09
24	−2.27	− .24	−1.26
25	−2.71	1.07	1.57
26	− .69	−1.59	1.39
27	1.96	.52	−1.86
28	.70	− .13	− .34
29	− .17	− .30	.25
30	− .22	1.28	.25
31	.10	− .16	− .01
32	−1.23	.15	− .42
33	.18	− .72	− .11
34	−1.42	− .65	− .51

NOTE: The Underlined values indicate extreme positive or negative coordinates for the items that define the dimension endpoints.

TABLE 3

Hierarchical Cluster Analysis Results
Based on Three-Dimensional Configuration

Cluster Number	Description	Items Included
I	Course Characteristics	5, 13, 17, 18, 19, 24, 25, 30, 32, 34
II	Lecture Style	1, 2, 16, 26
III	Instructor Attributes and Rapport with Students	3, 4, 6, 7, 8, 9, 10, 11, 12, 14, 15, 20 21, 22, 23, 27, 28, 29, 31, 33

REFERENCES

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1 - 27.

Kruskal, J. B., & Wish, M. Multidimensional scaling. Beverly Hills, Calif.: Sage Publications, Inc., 1978.

Kulik, J. A., & McKeachie, W. J. The evaluation of teachers in higher education. In F. N. Kerlinger (Ed.), Review of research in education (Vol. 3). Itaska, Ill.: F. E. Peacock Publishers, 1975.

Shepard, R. N., Romney, A. K., & Nerlove, S. B. Multidimensional scaling: theory and applications in the behavioral sciences (Vol. I). New York: Seminar Press, 1972.

Young, F. W., Takane, Y., & Lewyckyj, R. ALSCAL: a nonmetric multidimensional scaling program with several individual differences options. Journal of Marketing Research, 1978, 15, 612-615; and Behavioral Research Methods and Instrumentation, 1978, 10, 451-453.

APPENDIX A

THE UNIVERSITY OF TEXAS AT AUSTIN - COURSE-INSTRUCTOR SURVEY: EDUCATIONAL PSYCHOLOGY

DIRECTIONS: PLEASE MARK YOUR RESPONSES TO EACH ITEM ON THE SEPARATE ANSWER SHEET.

1. My classification is: A = Graduate B = Senior C = Junior D = Sophomore E = Freshman
2. My sex is: A = Male B = Female
3. My final grade in this course will probably be: A = A B = B C = C D = D E = F
4. My overall grade-point average at UT Austin is:

   A = 3.50-4.00 B = 3.00-3.49 C = 2.50-2.99 D = 2.00-2.49 E = Less than 2.00
5. The college or school in which I am enrolled is:

   A = Liberal Arts or Natural Sciences B = Business Administration C = Education
   D = Fine Arts E = Other
6. I took this course to satisfy: A = Major or minor field requirements
   B = Other specific degree requirements D = Non-degree requirements(e.g., teacher cer.)
   C = Elective credits required for degree E = No requirements at all

Items 7-36 all use the same response scale in which:

A = Very Satisfactory
B = Satisfactory
C = Barely Satisfactory
D = Unsatisfactory
E = Not applicable

If you do not feel the items describes an activity relevant to the course, mark the "E" position. If you feel the item is applicable, use "A" through "D" according to your opinion as to how well it was handled.

7. The instructor had a thorough knowledge of the subject matter.
8. Lectures were well-planned, organized and clearly presented.
9. The instructor was aware of whether or not the class members were following his/her discussion or lectures with understanding.
10. The instructor held my attention and interest during class.
11. This course has increased my ability to critically evaluate work in this field.
12. The instructor commented informatively on my written work, either orally or in writing.
13. The instructor explored different points of view and helped to increase my awareness of alternatives.
14. The instructor emphasized major points and made clear their relationships to one another.
15. The instructor made me feel free to ask questions, disagree and express my ideas.
16. The instructor was enthusiastic about the subject.
17. Taking this course has increased my skills in areas relevant to my future work.
18. The instructor was accessible to students outside of class.
19. The instructor related the subject matter to other areas of knowledge.
20. The examples and illustrations used in lecture and/or discussion made the material clearer to me.
21. The instructor, in his/her dealings with students, respected them as individuals.
22. The instructor was self-confident before the class.
23. This course has stimulated me to take other related courses or to read on my own.

CONTINUED ON THE BACK.

24. Assignments and reading were chosen to increase my understanding of the subject matter.
25. The instructor related current developments in the field to the course material.
26. The instructor paced the course well.
27. The instructor was sensitive to the feelings and needs of students.
28. The instructor was a dynamic and energetic person.
29. The instructor got me interested in the subject.
30. The instructor gave examinations that required creative, original thinking.
31. The instructor was able to suggest optional reading for further study of specific areas.
32. Lectures did not unduly overlap the assigned readings.
33. The instructor related class topics to students' lives and experiences.
34. The instructor showed a genuine concern with the quality of his/her teaching.
35. The instructor was intellectually stimulating.
36. The instructor has sufficient evidence to evaluate fairly my achievement in this course.

For Items 37-45, choose the appropriate response from those given for each item.

37. Given the opportunity, I would choose this instructor again for another course.

    A = Definitely yes B = Yes C = Uncertain or neutral D = No E = Definitely no

38. At this time, I feel this course will be (or has already been) of value to me.

    A = Definitely yes B = Yes C = Uncertain or neutral D = No E = Definitely no

39. Compared with all the instructors I have had in college and/or graduate school, this instructor was:

    A= One of the best B= Above average C= Average D= Below average E= Far below average

40. Compared with all the courses I have had in college and/or graduate school, this course was:

    A= One of the best B= Above average C= Average D= Below average E= Far below average

As an option for Items 41-45, the instructor has been asked to list up to five major objectives which he/she feels should have been attained in this course. Indicate the extent to which you feel you have mastered each of these objectives by using the appropriate response as follows:

A = Objective Understood and Completely Attained
B = Objective Understood and Largely Attained
C = Objective Understood and Partially Attained
D = Objective Understood, No Attainment
E = Objective Not Understood

PLEASE NOTE THESE INSTRUCTIONS FOR THE COMMENTS SECTION OF THE ANSWER SHEET. Instructors have indicated that written comments frequently contain very constructive recommendations and many see the comments as providing some of the most specifically helpful information to them. Please take the time and effort to make comments. If the instructor has listed objectives in items 41-45, it would be helpful if you would expand on your responses to those items. In addition, your instructor would like to know if there is something you believe he/she has done especially well in the teaching in this course.